Specification of Race and Ethnicity
Draft: January 3, 2005

Objective: To develop a common agreement on the expression of Race and Ethnicity.

Purpose: Agreements on formats and permissible values are especially important tools because they reduce the ambiguity between the same data elements residing in different information tools and enable the efficient and integrated flow of data. When all partners consistently use the same terminology and framework, data exchange and analysis is improved.

The purpose of the Race and Ethnicity standard is to define basic template data elements and permissible values, and to develop a consensus data standard for the exchange and analysis of Race and Ethnicity information.

Standard Description: The Race and Ethnicity standard will consist of data elements that describe the components of race information and will provide guidance for recording and exchanging this type of information. The following Candidate data standard is based on the United States Office of Management and Budget (OMB) for reporting federal information about Race and Ethnicity.

Definitions of Key Terms:

- Race – an arbitrary classification based on physical characteristics; a group of persons related by common descent or heredity.
- Ethnicity – an arbitrary classification based on cultural, religious, or linguistic traditions; ethnic traits, background, allegiance, or association.
- Display Format – The way characters are presented or arranged in a report or on a computer screen.
- Data Entry Format – The way characters are entered on a paper form or on a computer screen.
- Exchange Format – The way characters are presented or arranged in an electronic file used to transfer between applications.
- Storage Format – The way characters are recorded in a database or computer application.

Applicability: Information collected about Race and Ethnicity may be shared among applications and included in regulatory reports. An agreement on common ways of exchanging Race and Ethnicity information will facilitate this information exchange. The United States Office of Management and Budget (OMB) has directed that all Federal programs adopt the OMB Standards for Classification of Federal Data on Race and

Ethnicity no later than January 1, 2003. In no case shall the provisions of the standards be construed to limit the collection of data to the categories described. The collection of greater detail is encouraged; however, any collection that uses more detail shall be organized in such a way that the additional categories can be aggregated into these minimum categories for data on Race and Ethnicity. Appendix A of this document provides a draft summary for application of the OMB standard to clinical trials as prepared by the Food and Drug Administration, http://www.fda.gov/cder/guidance/5054dft.doc

Related Standards:

Summary of OMB Revisions to Standards for Classification of Federal Data on Race and Ethnicity, Federal Register Notice, October 30, 1997, http://www.whitehouse.gov/omb/fedreg/ombdir15.html

OMB developed this standard to provide flexibility and ensure data quality when reporting Race and Ethnicity. Information may be collected either as one question (aggregate) or as two questions (separate). The two-question format is preferred. The standard notes that when Race and Ethnicity are collected separately, Ethnicity shall be collected first. When data on Race and Ethnicity are collected separately, provision shall be made to report the number of respondents in each racial category who are Hispanic or Latino. When aggregate data are presented, data producers shall provide the number of respondents who marked (or selected) only one category, separately for each of the five racial categories. In addition to these numbers, data producers are strongly encouraged to provide the detailed distributions, including all possible combinations, of multiple responses to the race question. If data on multiple responses are collapsed, at a minimum the total number of respondents reporting "more than one race" shall be made available. Table 1 outlines the permissible values (provided in bold) and value meanings (where provided in the OMB documentation) for ethnicity and race categories that must be used for reporting information to OMB.

Table 1 OMB Specification of Data Elements for Ethnicity and Race

Data Category	**Permissible Values**
OMB Ethnicity Category	**Hispanic or Latino*** (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can also be used in addition to "Hispanic or Latino.") **Not Hispanic or Latino** (*No definition is provided in the OMB documentation.*)
OMB Race Category	**American Indian or Alaska Native** (A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.) **Asian** (A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.) **Black or African American (**A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American.") **Native Hawaiian or Other Pacific Islander** (A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.)

Data Category	Permissible Values
	White (A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.)

***** Because regional usage of the terms differs -- Hispanic is commonly used in the eastern portion of the United States, whereas Latino is commonly used in the western portion -- this change may contribute to improved response rates.

Proposed NCI/caBIG Data Elements:

To provide maximum flexibility and to ensure quality data, the OMB permits the collection of Race and Ethnicity data in either a two-question or single-question format. Only the two-question format has been included for consideration in this document. OMB allows the collection of more than one value for Race.

Table 2 lists the data elements are recommended for use when collecting Race and Ethnicity data for National Cancer Institute (NCI) and the Cancer BioInformatics Grid (caBIG). The proposed permissible values are based on the values specified in the OMB standard as outlined in Table 1. The values of Unknown and Not Reported have been added to the permissible value lists for each data element to support the information collection needs of current programs. The data elements are also provided for reporting Race or Ethnicity categories not included in the OBM list. OMB requires that any other Race or Ethnicity categories collected must be aggregated to the OMB categories when reporting federal data.

Table 2 NCI/caBIG Specification of Data Elements for Ethnicity and Race

Data Elements (CDE ID:Version)	Definition	Datatype (Max Number of Characters)	Permissible Values
Ethnicity Category Text (2192217:1)	The text for reporting information about ethnicity based on the Office of Management and Budget (OMB) categories.	Character (22)	**Hispanic or Latino*** (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can also be used in addition to "Hispanic or Latino.") **Not Hispanic or Latino** (A person not meeting the definition for Hispanic or Latino.) **Unknown** (Could not be determined or unsure.) **Not Reported** (Not provided or available.)
Race Category Text (2192199:1)	The text for reporting information about race based on the Office of Management and	Character (41)	**American Indian or Alaska Native** (A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.)

Data Elements (CDE ID:Version)	Definition	Datatype (Max Number of Characters)	Permissible Values
	Budget (OMB) categories.		**Asian** (A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.) **Black or African American (**A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American.") **Native Hawaiian or Other Pacific Islander** (A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.) **White** (A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.) **Unknown** (Could not be determined or unsure.) **Not Reported** (Not provided or available.)
Other Ethnicity Category Text (2192221:1)	The text that describes an ethnicity not included in the Office of Management and Budget (OMB) list for reporting ethnicity data.	Character (200)	
Other Race Category Text (2192205:1)	The text that describes a race not included in the Office of Management and Budget (OMB) list for reporting race data.	Character (200)	

*Because regional usage of the terms differs -- Hispanic is commonly used in the eastern portion of the United States, whereas Latino is commonly used in the western portion -- this change may contribute to improved response rates.

Specified code sets could be created for Race and Ethnicity, however this current document does not include code sets for either Race or Ethnicity. For example, the **Cancer Therapy Evaluation Program (CTEP) Clinical Data Update System (CDUS)** system uses the following codes for Ethnicity: 1 - Hispanic or Latino, 2 - Non-Hispanic, 8 - Not Reported, and 9 - Unknown. Similarly, CDUS uses the following codes for Race:

01- White, 03 Black or African American, 04 - Native Hawaiian or other Pacific Islander, 05 - Asian, 06 - American Indian or Alaska Native, 9 - Not Reported, and 99 – Unknown. The **Centers for Disease Control (CDC)** also has an extensive coded list for Race and Ethnicity (see URL provided in References). The CDC coded list is under consideration by HL7 for use in HL7 messages. Permissible values include a Unique Identifier and a Hierarchical Code. For Race the CDC coded list uses: 1000-9 (Unique Identifier) and R1 (Hierarchical Code)– American Indian or Alaska Native, 2028-8 and R2 – Asian, 2054-4 and R3 – Black or African American, 2076-8 and R4 – Native Hawaiian or other Pacific Islander2106-3 and R5 – White, and 2131-1 and R9 - Other Race. Detailed Hierarchical Codes with associated Unique Identifiers are provided for recording more detailed Race information. For Ethnicity, CDC uses 2135-2 and E1 – Hispanic or Latino, and 2186-5 and E2 – Not Hispanic or Latino.

Draft Age, Sex, and Race/Ethnicity guidelines developed by the **DCEP ITOC Subcommittee** identify these three variables that are of interest to all studies conducted in that division. The report summarizes information gathered from various studies in DCEG as well as from the Census Bureau and other outside sources. The results of their review for recording of Race and Ethnicity information is provided in Appendix B. These results illustrate the complexity of this issue, particularly for demographic data collection.

Issues Not Addressed in This Standard:

- Mapping of additional detailed race and ethnicity values to the specified OMB categories.
- Specific lists to meet programmatic needs to request addition Race and Ethnicity information using the "Other Ethnicity or Race" constructs.
- Coded values for Race and Ethnicity categories.

References:

- Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, Office of Management and Budget (OMB), Federal Register Notice, October 30, 1997, http://www.whitehouse.gov/omb/fedreg/ombdir15.html

- Food and Drug Administration (FDA) Issues Guidance for Collection of Race and Ethnicity Data in Clinical Trials for FDA Regulated Products, http://www.fda.gov/bbs/topics/ANSWERS/2003/ANS01193.html and http://www.fda.gov/fdac/features/2003/303_race.html

- National Institutes of Health (NIH) Policy on Reporting Race and Ethnicity Data: Subjects in Clinical Research, http://grants.nih.gov/grants/guide/notice-files/NOT-OD-01-053.html

- Center for Disease Control (CDC) Race and Ethnicity Code Set Version 1.0, http://www.cdc.gov/nedss/DataModels/Race_Ethnicity_CodeSet.pdf

- Cancer Therapy Evaluation Program (CTEP) Clinical Data Update System (CDUS), V3, Instructions and Guidelines, (may be accessed from CTEP home page, http://ctep.cancer.gov/, and choosing CDUS from the right menu.)

Appendix A

Draft Guidance for Industry
Collection of Race and Ethnicity Data in Clinical Trials
Prepared by the Food and Drug Administration

Draft Guidance for Industry

Collection of Race and Ethnicity Data in Clinical Trials

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to Dockets Management Branch (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (OC) Katherine Hollinger at (301)827-0350, (CDER) Nancy Derr at 301-594-5400, (CBER) Ilan Irony at 301-827-5378, or (CDRH) IDE Staff at (301)594-1190.

Office of Training and Communication
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm
or
Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm.
(Tel) Voice Information System at 800-835-4709 or 301-827-1800
or

Center for Devices and Radiological Health (HFZ-220)
Food and Drug Administration
1350 Piccard Drive
Rockville, MD 20850-4307 U.S.A.
http://www.fda.gov/cdrh/ggpmain.html
Email: dsma@cdrh.fda.gov
Fax: 301.443.8818
(Tel) Manufacturers Assistance: 800.638.2041 or 301.443.6597

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Office of the Commissioner (OC)

January 2003
Procedural

TABLE OF CONTENTS

Guidance for Industry[1]
Collection of Race and Ethnicity Data
in Clinical Trials

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statutes and regulations.

I. INTRODUCTION

This guidance recommends a standardized approach for collecting race and ethnicity information in clinical trials conducted in the United States and abroad for certain FDA regulated products. The standardized approach being recommended was developed by the Office of Management and Budget (OMB). The guidance explains the OMB categories and FDA's reasons for recommending the use of the OMB categories.[2]

This document provides guidance on the requirements set forth in the 1998 final rule on Investigational New Drug Applications and New Drug Applications[3] (Demographic Rule). The Demographic Rule requires sponsors of new drug applications (NDAs) to include in their applications analyses of effectiveness and safety data for important demographic subgroups, including racial subgroups.[4]

Although the regulations governing medical devices do not include requirements for the collection of demographic data comparable to those for INDs and NDAs described above, for those cases in which race and ethnicity data are relevant to determining the safety and effectiveness of a device, FDA encourages sponsors to collect the data in accordance with the OMB information collection standards discussed in this guidance document. Sponsors are also encouraged to discuss any race or ethnicity issue with the appropriate review division within the

[1] This guidance has been developed by the Race and Ethnicity Working Group from the Office of the Commissioner, the Center for Biologics Evaluation and Research (CBER) and the Center for Drug Evaluation and Research (CDER) of the Food and Drug Administration (FDA). This guidance is an FDA Data Council initiative to recommend the standardized collection of health information in a regulatory environment.

[2] The Agency is developing a companion guidance that will make recommendations about how to analyze and report race and ethnicity data.

[3] 63 FR 6854 (Feb. 11, 1998) (codified at 21 CFR 312.33(a)(2) and 21 CFR 314.50(d)(5)).

[4] 21 CFR 314.50(d)(5)(vi)(a).

Office of Device Evaluation, Center for Devices and Radiological Health, when developing their study protocols.

This guidance does not discuss increasing the number of studies in which subpopulations are exposed to a product. The guidance also does not discuss increasing the total number of participants or members of a subpopulation in clinical trials.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND

FDA regulations require sponsors to present in certain marketing applications an analysis of data according to demographic subgroups (age, gender, race), as well as an analysis of modifications of dose or dosage intervals for specific subgroups (21 CFR 314.50 (d)(5)(vi)(a)).[5]

In 1997, OMB issued its revised recommendations for the collection and use of race and ethnicity data by Federal agencies (Policy Directive 15).[6] FDA now recommends the use of the standardized OMB race and ethnicity categories for data collection in clinical trials for two reasons. First, the use of the recommended OMB categories will help ensure consistency in demographic subset analyses across studies used to support certain marketing applications to FDA and across data collected by other government agencies (21 CFR §§ 312.120 and 314.106 (b); see also *E5 Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data*[7]). Second, these categories may be useful in evaluating potential differences in the safety and efficacy of pharmaceutical products among population subgroups. To assess potential subgroup differences in a meaningful way, it is important to provide guidance on the use of uniform, standard categories in data collection for racial and ethnic subgroups.

[5] Under 21 CFR 314.101(d)(3), the Agency may refuse-to-file an NDA if there is an inadequate evaluation for safety and/or effectiveness of the population intended to use the drug, including pertinent subsets, such as gender, age, and racial subsets. See FDA Guidance to Industry *Refusal to File* July 1993.

[6] Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting, 1997 (reprinted in Appendix 2). See also OMB guidance entitled *Implementation of the 1997 Standards for Federal Data on Race and Ethnicity* (2000).

[7] This guidance was developed under the auspices of the International Conference of Harmonisation (ICH). Reprinted at 63 FR 31790 (June 10, 1999).

A. Relevance of Population Subgroup Studies

The OMB stated that its race and ethnicity categories were nonanthropologic (in other words, not scientifically based) designations but, instead, were categories that described the sociocultural construct of our society. The Department of Health and Human Services (HHS) chose to adopt these standardized categories for its agencies that report statistics, as the categories are relevant to assessing various health related data, including public health surveillance and research. FDA believes that the use of the OMB categories will facilitate comparisons across clinical studies analyzed by the FDA and with data collected by other agencies. Collection of data using standard categories may enhance patient safety by helping the Agency evaluate potential differences in drug response among subpopulations and may help facilitate analyses seeking differences in response.

Some differences in response to medical products have already been observed in racially and ethnically distinct subgroups of the U.S. population. These differences may be attributable to intrinsic factors (e.g., genetic, clearance[8]), extrinsic factors (e.g., diet, environmental exposure, sociocultural issues), or interactions between these factors. For example, in the United States, Whites[9] are more likely than persons of Asian and African heritage to have abnormally low levels of an important enzyme (CYP2D6) that metabolizes drugs belonging to a variety of therapeutic areas, such as antidepressants, antipsychotics, and beta blockers (Xie 2001). Additionally, after using some drugs in the psychotherapeutic class, slower enzyme metabolism (CYP2C19) has been observed in persons in the United States of Asian descent as compared to Whites and Blacks (Xie 2001). Other studies have shown that Blacks respond poorly to several classes of antihypertensive agents (beta blockers and angiotensin converting enzyme (ACE) inhibitors) (Exner 2001 and Yancy 2001). Racial differences in skin structure and physiology have been noted that can affect response to dermatologic and topically applied products (Taylor 2002). Clinical trials have demonstrated lower responses to interferon-alpha used in the treatment of hepatitis C among Blacks when compared to other racial subgroups (McHutchison 2000 and Reddy 1999).

Collecting race and ethnicity data using standardized categories will enhance the early identification of differences in physiological response among racial and ethnic subgroups during the evaluation of safety and effectiveness of FDA-regulated products.[10] Furthermore, collection

[8] Clearance is a measure of drug or biologic elimination from the body.

[9] The terms used in this guidance to describe the various racial and ethnic groups are those used by OMB.

[10] The importance of understanding population subgroup differences for effective risk management is discussed in greater detail in various reports, the published scientific literature, and Agency guidance documents and regulations (see the bibliography for a listing of relevant documents).

of this data using standardized categories will facilitate comparisons across studies analyzed by FDA and with data collected by other Federal agencies.

B. FDA Decision to Recommend Use of the OMB Categories

Although the FDA has long requested race and ethnicity data on subjects in certain clinical trials, the Agency has not made explicit recommendations on the categories to use when collecting and reporting the data. In 1998, the Agency issued the Demographic Rule, which reflected the growing recognition within the Agency and the health community that (1) different subgroups of the population may respond differently to a specific drug product, and (2) although the effort should be made to look for differences in effectiveness and adverse reactions among such subgroups, that effort is not being made consistently.[11] In the Demographic Rule, the Agency discussed the importance of collecting data in clinical trials (and of analyzing and presenting those data in applications to the Agency) on population subgroups organized by gender, race, age, and other relevant subgroups. The Agency recommended that sponsors ask subjects in certain clinical trials to identify their racial group and, if desired, to use the OMB categories when collecting race and ethnicity data.[12]

During the past two decades, efforts have been under way in a number of Federal organizations to collect race and ethnicity data in Federal programs in a standardized way. (See Appendix 1 for a summary of those efforts). In 1997, HHS issued a statement entitled *Policy Statement on Inclusion of Race and Ethnicity in DHHS Data Collection Activities.* In this policy statement, HHS adopted the revised OMB categories for including race and ethnicity in HHS-funded and -sponsored data collection and reporting systems. The HHS policy states that the categories described in the revised OMB Directive 15 and its future revisions should be used when collecting and reporting data in HHS data systems or reporting HHS-funded statistics.

The collection of standardized race and ethnicity data across HHS improves the quality and consistency of reported health statistics. To be consistent with HHS policy and to facilitate and enhance data consistency and comparability, FDA has decided to recommend the OMB Directive 15 categories be used for the collection of race and ethnicity data in clinical trials.[13] The Agency recommends that sponsors use the categories as outlined in this guidance when

[11] 63 FR 6855 (1998).

[12] In the preamble to the final rule, FDA stated that it did not believe it was necessary to define specific racial categories in the rule itself because drug sponsors have been successful in identifying the relevant racial categories to examine safety and efficacy profiles of drugs (63 FR 6859). However, the FDA now believes that using uniform categories will enhance the consistency and comparability of data across studies submitted in marketing applications and across other government reported statistics.

[13] OMB directed these activities to begin by January 1, 2003, in all Federal programs, including HHS. Although FDA sought and received a variance from OMB exempting the Agency from reporting data using the Directive 15 categories in the past, FDA now recommends the use of the categories to enhance data consistency.

collecting race and ethnicity data in clinical studies for FDA-regulated products conducted in United States and abroad. More detailed race and ethnicity data can be collected when appropriate to the study or locale, but we recommend that the OMB categories be identified for all clinical trial participants when submitting such data to the Agency.

III. COLLECTING RACE AND ETHNICITY DATA IN CLINICAL TRIALS

The Agency recommends sponsors collect race and ethnicity data for clinical study participants as specified in the OMB Directive and its revisions. However, the recommendations in this section reflect the Agency's interest in more consistency in data collection. The Agency recommends the use of a two-question format, and to have trial participants self-report their racial and ethnic category to enhance consistency in the collection of the reported information. Based on the current OMB Directive, the Agency provides the following recommendations for the collection of the data:

1. We recommend using the two-question format for requesting race and ethnicity information, with the ethnicity question preceding the question about race.

2. We recommend that study participants self-report race and ethnicity information whenever feasible, and individuals be permitted to designate a multiracial identity. When the collection of self-reported designations is infeasible (e.g., because of the subject's inability to respond), we recommend the information be requested from a first-degree relative or other knowledgeable source.

3. For ethnicity, we recommend the following minimum choices be offered:
   - Hispanic or Latino
   - Not Hispanic or Latino

4. When race and ethnicity information is collected separately, we recommend the following minimum choices be offered for race:
   - American Indian or Alaska Native
   - Asian
   - Black or African American
   - Native Hawaiian or Other Pacific Islander
   - White

5. In certain situations, as recommended in OMB Directive 15, more detailed race and ethnicity information may be desired (e.g., *White* can reflect origins in Europe, the Middle East, or North Africa; *Asian* can reflect origins from areas ranging from India to Japan). If more detailed characterizations of race or ethnicity are collected to enhance data consistency, we recommend these characterizations be traceable to the five minimum designations for race and two designations for ethnicity listed in numbers 3 and 4.

IV. Clinical Trials Conducted Outside of the United States

To assist in assessing the relevance of foreign study population data to United States populations, we recommend sponsors use the OMB standardized categories when collecting data from study participants in clinical trials conducted outside of the U.S. However, FDA recognizes that the recommended categories for race and ethnicity were developed in the United States and that these categories may not adequately describe racial and ethnic groups in foreign countries. Therefore, we recommend collecting the data using more detailed categories of race and ethnicity to provide sponsors the flexibility to adequately characterize race and ethnicity for studies conducted outside the United States. If sponsors choose to use more detailed characterizations of race and ethnicity, it is important for analytical purposes that the data trace back to the recommended categories described below.

A. For ethnicity, we recommend that collection of data in foreign countries follow the same methods and designations used in the collection of data in U.S. clinical studies:

- Hispanic or Latino
- Not Hispanic or Latino

B. For racial designations in clinical studies conducted in foreign countries, we recommend that the categories be modified to reflect the following as appropriate:

- American Indian or Alaska Native
- Asian
- Black, of African heritage
- Native Hawaiian or Other Pacific Islander
- White

Note that the categories for inside and outside the US are very similar — only the description of the Black designation varies from the OMB recommended categories.

Appendix 1: History of federal efforts In DATA collection on Race and ethnicity and other subpopulations

During the past 20 or more years, a number of U.S. government initiatives have tried to address questions related to whether to and how to collect race and ethnicity data. Major initiatives are reviewed briefly here.

Office of Management and Budget (OMB) Initiatives

In May 1977, the OMB, issued "Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting." The standards were developed in response to needs related to enforcing civil rights laws in education. These classifications were not to be interpreted as being scientific or anthropological in nature, nor viewed as determinants of eligibility for participation in any Federal program. They were developed in response to needs expressed by both the Executive Branch and the Congress to provide for the collection and use of compatible, non-duplicated, exchangeable race and ethnicity data by Federal agencies. This Directive specified four categorizations for race and two for ethnicity:

- American Indian or Alaskan Native
- Asian or Pacific Islander
- Black
- White

- Hispanic
- Not of Hispanic origin

The OMB Directive specified two questionnaire formats for data collection: (1) a format combining race and ethnicity and (2) a preferred format with two separate questions for race and ethnicity.

Since 1993, efforts have been under way to standardize the collection of race and ethnicity data to foster comparability across data collection and reporting systems. In 1997, revisions to OMB Directive 15, **"**Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" (see appendix 2), were published. These revisions specified the minimum racial and ethnic diversity categories to be used when race and ethnicity are included in data collection and reporting for Federal programs. The Directive does not require that race and ethnicity be included in data collection and reporting, rather, it specifies what formats and categories to use when collecting this kind of data.

The revised OMB standards contain the following changes:

- Introduce the option to report more than one race for multiracial persons
- Break the Asian or Pacific Islander category into two — one labeled Asian, the other Native Hawaiian or Other Pacific Islander
- Change Hispanic to Hispanic or Latino
- Change Black to Black or African American
- Strongly encourage the use of self-identification
- Maintain the two question format for race and Hispanic ethnicity, when self-identification is used (the Hispanic origin question should precede the race question)

The revised categorizations were described in an OMB guidance*, Implementation of the 1997 Standards for Federal Data on Race and Ethnicity* (2000), as sociopolitical and intended for use in the collection of health data among other types of statistics.

Department of Health and Human Services Initiatives

In 1999, the Department of Health and Human Services (HHS) issued a report, *Improving the Collection and Use of Racial and Ethnic Data in HHS*. The report describes HHS policy on collecting and reporting data on race and ethnicity for HHS programs. The HHS report asks for the inclusion of race and ethnicity categories in HHS-funded and -sponsored data collection and reporting systems in all HHS programs, including in both health and human and social services. This policy clearly states that the minimum standard categories in OMB Directive 15 and revisions should be used when collecting and reporting data in HHS data systems or reporting HHS-funded statistics. The policy was developed to (1) help monitor HHS programs, (2) determine that Federal funds are being used in a nondiscriminatory manner, and (3) promote the availability of standard race and ethnicity data across various agencies to facilitate HHS responses to major health and human services issues.

National Institutes of Health Initiatives

In 1993, the National Institutes of Health (NIH) Revitalization Act directed the NIH to establish guidelines for including women and minorities in NIH-sponsored clinical research. NIH was directed to ensure that, when women and minorities were included as subjects, trials were designed and carried out in a manner that would provide a valid analysis of any differences between study subgroups, specifically women, minorities, and other participants in the trial. NIH guidelines direct investigators to consider appropriate representation of subjects of different genders and racial and ethnic backgrounds to provide the opportunity for detecting major qualitative differences and to identify more subtle differences that may be explored more fully in specifically targeted studies.

NIH guidelines established that prior to the design of phase 3 studies, there must be a review of evidence to show whether clinically important gender and minority based differences are expected. This evidence may include, but is not limited to, data derived from prior animal studies, clinical observations, metabolic studies, genetic studies, pharmacology studies, and observational, natural history, epidemiology, and other relevant studies. For example, if men and women were thought to respond differently to an intervention, the phase 3 clinical trial must be designed to answer two separate primary questions, one for men and the other for women, with

adequate sample size for each. When prior studies neither support nor negate significant differences of clinical or public health importance, the phase 3 trial will be required to support the sufficient and appropriate accrual of participants by gender and race/ethnicity, so that a valid analysis of the intervention effects can be performed. However, the trial is not required to provide high statistical power for these comparisons. The term *valid analysis* refers generally to a reasonable descriptive approach to the data.

Food and Drug Administration Initiatives

Beginning in the 1980s, FDA grew concerned about possible differences in drug safety and efficacy among different population subgroups. Because the origins of subpopulation issues stem from the identification of differences in response in women and geriatric populations, references to those initiatives are included below. In 1983, the Agency initiated development of guidance on the study of drugs to be used in geriatric patients. The guidance *Guideline for the Study of Drugs Likely to be Used in the Elderly* was issued in 1989.

The first regulation specifying the analysis of population subsets appeared in 1985 in 21 CFR 314.50, which called for evidence to support the "dosage and administration section of the labeling, including support for the dosage and dose interval recommended, and modifications for specific subgroups (e.g., pediatrics, geriatrics, patients with renal failure..." 21 CFR 314.50(d)(5)).

In 1988, the Agency issued guidance describing elements of a New Drug Application's analysis of clinical study data. The *Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications* emphasized the importance of conducting subset analyses on data from clinical studies submitted in New Drug Applications (NDAs). This guidance specified race and ethnicity as types of population subsets for which separate analyses of data from clinical studies should be conducted for assessments of product safety and effectiveness.

In July 1993, FDA published a guidance on the study of drugs in both genders entitled *Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs*. The guidance specifically called for analysis of trials by gender and for evaluating pharmacokinetics in women. In the Federal Register notice announcing the guidance, FDA also abandoned the policy explained in a 1977 guidance excluding women of childbearing potential from participation in the earliest phases of clinical trials.

In 1993, FDA also published guidance on the Agency's use of the refusal-to-file (RTF) option if certain analyses were not performed. The guidance states that the Agency can exercise its RTF authority under 21 CFR 314.101(d)(3) if there is "inadequate evaluation for safety and/or effectiveness of the population intended to use the drug, including pertinent subsets, such as gender, age, and racial subsets."

In the Food and Drug Administration Modernization Act of 1997 (FDAMA), Congress directed FDA to examine issues related to the inclusion of racial and ethnic groups in clinical trials of

new drugs. Section 115(b) of FDAMA required the Secretary, "in consultation with the Director of the National Institutes of Health and with representatives of the drug manufacturing industry, [to] review and develop guidance, as appropriate, on the inclusion of women and minorities in clinical trials...." (codified at 21 USC 355(b)(1)). In response, FDA established the FDAMA Women and Minorities Working Group to review and implement this section of FDAMA. In a report issued on July 20, 1998, the Working Group concluded that the Agency would implement procedures to enhance its ability to gather and evaluate demographic data and then decide whether additional guidance should be developed in the future.

In 1998, the Agency published the Demographic Rule, which amended the language in 21 CFR § §312.33(a)(2) and 314.50(d)(5) requiring sponsors to tabulate the numbers of participants in clinical trials by age group, gender, and race in investigational new drug application (IND) annual reports and characterize the data in NDAs according to these same subgroups and, when appropriate, present safety data from other subgroups of the population of patients, such as for patients with renal failure or patients with different levels of severity of the disease.

In 1999, a guidance for industry entitled *Population Pharmacokinetics* made recommendations on the use of population pharmacokinetics in the drug development process to help identify differences in drug safety and efficacy among population subgroups, including race and ethnicity. This guidance recommends that industry conduct clinical studies in subjects representative of the population to be treated with the drug.

In 2000, the guidance *Content and Format of the Adverse Reactions Section of Labeling for Human Prescription Drugs and Biologics* was issued to assist sponsors in developing the adverse reactions section of labeling for human prescription drug and biological products. It recommended presentation of relevant racial and ethnicity subgroup information on product labeling.

The May 2001 guidance for industry *Clinical Studies Section of Labeling for Prescription Drugs and Biologics — Content and Format* explains that the clinical studies section of the labeling should include a summary statement about the results of the required explorations of treatment effects in age, gender, and racial subgroups. In October 2001, a guidance for industry was issued (*Content and Format for Geriatric Labeling*) that provides information on submitting geriatric (persons aged 65 years or older) labeling for human prescription drugs and biological products.

In 2002 in the Best Pharmaceuticals for Children Act, the FDA was directed to monitor the racial and ethnic designations of children participating in clinical studies for pharmaceutical products (Pub. L. 107-109, Jan. 4, 2002).

ICH E5 - Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data

In 1999, as part of an international effort among Japan, the European Union, and the United States to harmonize technical requirements for pharmaceutical drug development and regulation (ICH), the FDA published a guidance entitled *E5 Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data* to permit the clinical data collected in one region to be used in the registration or approval of a drug or biological product in another region, while

allowing for the influence of ethnic factors (63 FR 31790, June 10, 1999). The *E5* guidance defines ethnic factors that affect response in terms of both intrinsic and extrinsic issues. Because differences in ethnic factors have the potential to adversely affect some subpopulations, the *E5* guidance provides a general framework for how to evaluate medicines with regard to their sensitivity to ethnic factors.

Appendix 2: Revised Directive 15

OMB Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity

(Adopted on October 30, 1997)

This classification provides a minimum standard for maintaining, collecting, and presenting data on race and ethnicity for all Federal reporting purposes. The categories in this classification are social-political constructs and should not be interpreted as being scientific or anthropological in nature. They are not to be used as determinants of eligibility for participation in any Federal program. The standards have been developed to provide a common language for uniformity and comparability in the collection and use of data on race and ethnicity by Federal agencies.

The standards have five categories for data on race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. There are two categories for data on ethnicity: "Hispanic or Latino," and "Not Hispanic or Latino."

1. Categories and Definitions

The minimum categories for data on race and ethnicity for Federal statistics, program administrative reporting, and civil rights compliance reporting are defined as follows:

-- **American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

-- **Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

-- **Black or African American.** A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."

-- **Hispanic or Latino.** A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin," can be used in addition to "Hispanic or Latino."

-- **Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

-- **White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

Respondents shall be offered the option of selecting one or more racial designations. Recommended forms for the instruction accompanying the multiple response question are "Mark one or more" and "Select one or more."

2. Data Formats

The standards provide two formats that may be used for data on race and ethnicity. Self-reporting or self-identification using two separate questions is the preferred method for collecting data on

race and ethnicity. In situations where self-reporting is not practicable or feasible, the combined format may be used.

In no case shall the provisions of the standards be construed to limit the collection of data to the categories described above. The collection of greater detail is encouraged; however, any collection that uses more detail shall be organized in such a way that the additional categories can be aggregated into these minimum categories for data on race and ethnicity.

With respect to tabulation, the procedures used by Federal agencies shall result in the production of as much detailed information on race and ethnicity as possible. However, Federal agencies shall not present data on detailed categories if doing so would compromise data quality or confidentiality standards.

a. Two-question format

To provide flexibility and ensure data quality, separate questions shall be used wherever feasible for reporting race and ethnicity. When race and ethnicity are collected separately, ethnicity shall be collected first. If race and ethnicity are collected separately, the minimum designations are:
Race:
-- American Indian or Alaska Native
-- Asian
-- Black or African American
-- Native Hawaiian or Other Pacific Islander
-- White
Ethnicity:
-- Hispanic or Latino
-- Not Hispanic or Latino

When data on race and ethnicity are collected separately, provision shall be made to report the number of respondents in each racial category who are Hispanic or Latino.
When aggregate data are presented, data producers shall provide the number of respondents who marked (or selected) only one category, separately for each of the five racial categories. In addition to these numbers, data producers are strongly encouraged to provide the detailed distributions, including all possible combinations, of multiple responses to the race question. If data on multiple responses are collapsed, at a minimum the total number of respondents reporting "more than one race" shall be made available.

b. Combined format

The combined format may be used, if necessary, for observer-collected data on race and ethnicity. Both race (including multiple responses) and ethnicity shall be collected when appropriate and feasible, although the selection of one category in the combined format is acceptable. If a combined format is used, there are six minimum categories:
-- American Indian or Alaska Native
-- Asian
-- Black or African American

-- Hispanic or Latino
-- Native Hawaiian or Other Pacific Islander
-- White

When aggregate data are presented, data producers shall provide the number of respondents who marked (or selected) only one category, separately for each of the six categories. In addition to these numbers, data producers are strongly encouraged to provide the detailed distributions, including all possible combinations, of multiple responses. In cases where data on multiple responses are collapsed, the total number of respondents reporting "Hispanic or Latino and one or more races" and the total number of respondents reporting "more than one race" (regardless of ethnicity) shall be provided.

3. Use of the Standards for Record Keeping and Reporting

The minimum standard categories shall be used for reporting as follows:

a. Statistical reporting

These standards shall be used at a minimum for all federally sponsored statistical data collections that include data on race and/or ethnicity, except when the collection involves a sample of such size that the data on the smaller categories would be unreliable, or when the collection effort focuses on a specific racial or ethnic group. Any other variation will have to be specifically authorized by the OMB through the information collection clearance process. In those cases where the data collection is not subject to the information collection clearance process, a direct request for a variance shall be made to OMB.

b. General program administrative and grant reporting

These standards shall be used for all Federal administrative reporting or record keeping requirements that include data on race and ethnicity. Agencies that cannot follow these standards must request a variance from OMB. Variances will be considered if the agency can demonstrate that it is not reasonable for the primary reporter to determine racial or ethnic background in terms of the specified categories, that determination of racial or ethnic background is not critical to the administration of the program in question, or that the specific program is directed to only one or a limited number of racial or ethnic groups.

c. Civil rights and other compliance reporting

These standards shall be used by all Federal agencies in either the separate or combined format for civil rights and other compliance reporting from the public and private sectors and all levels of government. Any variation requiring less detailed data or data which cannot be aggregated into the basic categories must be specifically approved by OMB for executive agencies. More detailed reporting which can be aggregated to the basic categories may be used at the agencies' discretion.

4. Presentation of Data on Race and Ethnicity

Displays of statistical, administrative, and compliance data on race and ethnicity shall use the categories listed above. The term "nonwhite" is not acceptable for use in the presentation of Federal Government data. It shall not be used in any publication or in the text of any report. In cases where the standard categories are considered inappropriate for presentation of data on particular programs or for particular regional areas, the sponsoring agency may use:

a. The designations "Black or African American and Other Races" or "All Other Races" as collective descriptions of minority races when the most summary distinction between the majority and minority races is appropriate;

b. The designations "White," "Black or African American," and "All Other Races" when the distinction among the majority race, the principal minority race, and other races is appropriate; or

c. The designation of a particular minority race or races, and the inclusion of "Whites" with "All Other Races" when such a collective description is appropriate. In displaying detailed information that represents a combination of race and ethnicity, the description of the data being displayed shall clearly indicate that both bases of classification are being used.

When the primary focus of a report is on two or more specific identifiable groups in the population, one or more of which is racial or ethnic, it is acceptable to display data for each of the particular groups separately and to describe data relating to the remainder of the population by an appropriate collective description.

5. Effective Date

The provisions of these standards are effective immediately for all new and revised record keeping or reporting requirements that include racial and/or ethnic information. All existing record keeping or reporting requirements shall be made consistent with these standards at the time they are submitted for extension, or not later than January 1, 2003.

References

Exner, D., D. Dries, M. Donamski, and J. Cohn, 2001, "Lesser Response to Angiotensin-Converting-Enzyme Inhibitor Therapy in Black as Compared with White Patients with Left Ventricular Dysfunction," *N Engl J Med*, 344: 1351-1357.

McHutchison, J., T. Poynard, S. Pianko, S. Gordon, A. Reid, et. al., 2000, "The Impact of Interferon Alpha Plus Ribavirin on Response to Therapy in Black Patients with Chronic Hepatitis C. The International Hepatitis Interventional Therapy Group," *Gastroenterology*, 119: 1317-23.

Reddy, K. R., J. Hoofnagle, M. Tong, W. Lee, P. Pockros, et al., 1999, "Racial Differences in Responses to Therapy with Interferon in Chronic Hepatitis C. Consensus Interferon Study Group." *Hepatology*, 30: 787-93.

Taylor, S., 2002, "Skin of Color: Biology, Structure, Function, and Implications for Dermatologic Disease," *Journal of American Academy of Dermatology*, 46:S41-62.

Xie, H., R. Kim, A. Wood, and C. Stein, 2001,"Molecular Basis of Ethnic Differences in Drug Disposition and Response," *Annu Rev Pharmacol Toxicol*; 41:815-50.

Yancy, A., M. Fowler, W. Colucci, E. Gilbert, M. Bristow, et. al., 2001 "Race and the Response to Adrenergic Blockade with Cardefilol in Patients with Chronic Heart Failure," *N Engl J Med*, 344: 1358-1365.

BIBLIOGRAPHY

DHHS Policy and Reports

Department of Health and Human Services (DHHS), 1997, *Policy For Improving Race and Ethnicity Data, October 24, 1997,* http://www.hhs.gov/oirm/infocollect/nclusion.html.

DHHS, 1999, *Improving the Collection and Use of Racial and Ethnic Data in HHS*, joint report of the HHS Data Council Working Group on Racial and Ethnic Data and the Data Work Group of the HHS Initiative to Eliminate Racial and Ethnic Disparities in Health, December 1999, http://aspe.hhs.gov/datacncl/racerpt/index.htm.

FDA Regulations

Food and Drug Administration (FDA), 1998, "Investigational New Drug Applications and New Drug Applications," Final Rule, *Federal Register*, Vol. 63 p. 6854, February 11, 1998.

FDA, *Content and Format for Human Prescription Drugs; Addition of "Geriatric Use" Subsection in the Labeling*, Final Rule, *Federal Register*, Vol. 62, p. 45313, August 27, 1997.

FDA Guidance for Industry

FDA*, Clinical Studies Section of Labeling for Prescription Drugs and Biologics*, July 2001.

FDA, *Content and Format for Geriatric Labeling*, October 2001.FDA, *Content and Format of the Adverse Reactions Section of Labeling for Human Prescription Drugs and Biologics* (draft), June 2000.

FDA, *General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products* (draft), November 1998.

FDA, *General Considerations for the Clinical Evaluation of Drugs*, 1977.

FDA, *Format and Content of the Clinical and Statistical Sections of an Application*, July 1988.

FDA, *Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs*, July 1993.

FDA, *Study of Drugs Likely to be Used in the Elderly*, November 1989.

FDA, *In Vivo Drug Metabolism/Drug Interaction Studies: Study Design, Data Analysis, and Recommendations for Dosing and Labeling*, November 999.

FDA, Pharmacokinetics in Patients with Hepatic Insufficiency: Study Design, Data Analysis, and Impact on Dosing and Labeling (draft), December 7, 1999.

FDA, *Pharmacokinetics in Patients with Impaired Renal Function: Study Design, Data Analysis and Impact on Dosing and Labeling*, May 1998.

FDA, *Population Pharmacokinetics*, February 1999.

FDA, *Refusal to File*, July 12, 1993.

ICH Guidances

International Conference on Harmonisation (ICH), 2000, *E11 Clinical Investigation of Medicinal Products in the Pediatric Population*, ICH, December 2000.

ICH, *E4 Dose Response Information to Support Drug Registration*, ICH, November 1994.

ICH, *E7 Studies in Support of Special Populations: Geriatrics*, ICH, August 1994.

ICH, *E5 Ethnic Factors in the Acceptability Of Foreign Data*, ICH, June 1998.

Other Sources

Office of Management and Budget, 1994, Standards for the Classification of Federal Data on Race and Ethnicity, http://www.whitehouse.gov/omb/fedreg/notice_15.html.

General Accounting Office, *FDA Needs to Ensure More Study of Gender Differences in Prescription Drug Testing*, GAO/HFD-93-17.

NIH Revitalization Act of 1993, (PL 103-43).

NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research- Updated August 1, 2000. available at http://grants.nih.gov/grants/funding/women_min/women_min.htm.

Food and Drug Administration Modernization Act of 1997 (FDAMA) (P.L. 105-115).

Best Pharmaceuticals Act for Children of 2002 (P.L.107-109).

Appendix B

Extract from ITOC Subcommittee Common Data Elements Implementation Guide for Specification of Race and Ethnicity (Draft)

RACE/ETHNICITY

Questions about race and ethnicity are central to many DCEG studies. They are also considered to be among the most sensitive of questions to pose. For studies in the United States, the Office of Management and Budget has proposed a particular format for questions about race and ethnicity. The questions on this topic will vary depending upon the population in which the study is to be carried out.

Points to be considered, particularly for U.S. studies:

- how the data was obtained (self-report, external source)
- race and ethnicity (Hispanic/non) separate or combined (on the questionnaire or for the analysis)
- handling more than one reported race
- handling changes in an individual's self-categorization over time
- coding categories
  - 2000 Census codes and descriptions are detailed at http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf
  - coding for race varies by study design and the origin of the data
    - White, Black/African American, Asian, American Indian/Alaska native, Native Hawaiian/Pacific Islander, Other
    - White, Black, Other
    - >1 race allowed, combined variable: (e.g., only White, any Black, any Asian, Other)
  - Hispanic is commonly coded as a separate yes/no
  - a combination of race and ethnicity
    - White Hispanic, White non-Hispanic, Black Hispanic, Black non-Hispanic, Other

Examples of Race and Ethnicity Questions for Various Sources:

1. Office of Management and Budget (OMB) ethnicity questions: short form

1. Do you consider yourself to be:
   Hispanic or Latino
   Not Hispanic or Latino
2. What is your race? (Show Card)
   American Indian or Alaska Native
   Asian
   Black or African American
   Native Hawaiian or Other Pacific Islander
   White

2. ETHNICITY QUESTIONS: LONG FORM

MOST PEOPLE IN THE UNITED STATES HAVE ANCESTORS WHO CAME FROM OTHER PARTS OF THE WORLD. SOME PEOPLE HAVE MIXED ETHNIC BACKGROUNDS.

A13. What is your father's ethnic background?	A14. What is your mother's ethnic background?

RECORD BELOW. IF MORE THAN ONE ETHNICITY IS GIVEN FOR THE FATHER OR MOTHER, PROBE FOR PRIMARY ETHNICITY. RECORD BOTH PRIMARY AND SECONDARY ETHNICITIES.

Father's

	Primary	Secondary		Primary	Secondary
English, Scotch, Welsh	01	01	Central American	16	16
French	02	02	French Canadian	17	17
German	03	03	Mexican	18	18
Greek	04	04	Puerto Rican	19	19
Irish	05	05	South American	20	20
Italian	06	06	West Indian	21	21
Spanish, Portuguese	07	07	Chinese	22	22
Other European	08	08	Indian, Pakistani	23	23
Czechoslovakian	09	09	Japanese	24	24
Russian	10	10	Other Asian Countries or Pacific Islanders	25	25
Other Eastern European (Polish, Lithuanian, etc)	11	11	African	26	26
Swedish	12	12	Middle Eastern	27	27
Other Scandinavian (Norwegian, Danish, Finnish)	13	13	Other _____ (SPECIFY)	97	97
American Indian	14	14	Unknown	99	99
Canadian (Non-French)	15	15	N/A	00	00

Mother's

Note: The choices given here are the same as above.

3. Proposed Question in a Study of Breast Cancer in Egypt

Ethnicity:

1 Egyptian
2 Nubian
3 Some other European ethnicity

4. Questionnaire Developed in the Early 1980s

Which of these best describes your race or ethnic background?

01 White (Not Hispanic)	05 American Indian/Eskimo
02 White, Hispanic background	06 Japanese
03 Black (Not Hispanic)	07 Chinese
04 Black, Hispanic background	08 Other Oriental or Asian
	09 Other (specify)

5. Gansu Lung Cancer Study

Study Subject's Nationality: HAN……
OTHER…….
SPECIFY……..

6. Inherited Bone Marrow Failure Study

The race question is asked first and then ethnicity is asked a few questions later.

A2. Do you consider yourself to be:

Hispanic or Latino..... 1
Not Hispanic or Latino..... 2

A3. What is your race? (CHECK ALL THAT APPLY)

American Indian or Alaska Native..... 01
Asian..... 02
Black or African American..... 03
Native Hawaiian or Other Pacific Islander... 04
White..... 05

A4. Are you adopted?

YES..... 1
NO..... 2

A5. Where were you born?

_____

City State/Province Country

DON'T KNOW..... 8

A9. Where was your biological mother born?

_____

City State/Province Country

DON'T KNOW..... 8

A10. Where was your biological father born?

_____

City State/Province Country

DON'T KNOW..... 8

MOST PEOPLE HAVE ANCESTORS WHO COME FROM OTHER PARTS OF THE WORLD:

A11. What is your biological mother's ancestral background? (CHECK ALL THAT APPLY)

01 African	15 Hungarian	29 Puerto Rican
02 Canadian (French)	16 Indian (subcontinent)	30 Russian
03 Canadian (English)	17 Irish	31 Scottish
04 Central American	18 Italian	32 Scotch Irish
05 Chinese	19 Japanese	33 South American
06 Cuban	20 Korean	34 Spanish
07 Czechoslovakian	21 Mexican	35 Swedish
08 Danish	22 Middle Eastern	36 Turkish
09 Dutch	23 Native American	37 Ukrainian
10 English	24 Norwegian	38 Vietnamese
11 Finnish	25 Pacific Islander	39 Welsh
12 French	26 Pakistani	40 West Indian
13 German	27 Polish	96 Other (SPECIFY)_____
14 Greek	28 Portuguese	98 DON'T KNOW

Note: Question A12 for the biological mother's ancestral background gives the same choices.

CONCLUSIONS

Our examination of these three apparently simple, straightforward variables illustrates the difficulties inherent in providing any kind of standardized questionnaire element as well as the types of modifications that can make such elements suitable for a range of studies.